U.S. Geological Survey (USGS)
Community for Data Integration (CDI)
Request for Proposals (RFP)

For Fiscal Year 2018

Issue Date: December 18, 2017

Table of Contents

Overview

This document describes the CDI Request for Proposals (RFP) process for fiscal year 2018 (FY18). The CDI RFP consists of a two-stage process: Statements of Interest (SOI) and Full Proposals. Contact *cdi@usgs.gov* with any questions about the CDI RFP. Additional resources for the RFP are available on the proposals section of the CDI wiki: *https://my.usgs.gov/confluence/x/koReIQ*.

New in FY18

- The FY18 CDI RFP has a topical emphasis and encourages submissions that support the USGS Integrated Predictive Science Capacity in the specific focus area of risk assessment and hazard-exposure. A portion of the funds will be dedicated to this topic. However, all submissions that follow the CDI Guiding Principles will be considered. See more information in the "FY18 Topical emphasis" section below.
- The Statement of Interest narrative is a maximum of 1 page. (In FY17 it could be up to 1.5 pages.)
- In Phase 1 of the RFP, after the deadline for submission of Statements of Interest, an online session for 1-minute lightning talks will be held to help the CDI community to learn about the proposed ideas in preparation for commenting and voting.
- This year, the CDI will distribute funds to the lead USGS Science Center only, and the lead Science Center will be in charge of further distributions.
- A new, streamlined Data Management Plan form is provided for the Full Proposal stage.

Eligibility

Any USGS Mission Area, Region, Program, Center, Office, or duty station and their partner(s) are encouraged to apply. All proposals must specify a **USGS Federal employee** as lead Principal Investigator (PI). USGS personnel may be involved in more than one prospective or existing proposal, but may be the lead PI on only *one* proposal.

Estimated Available Funds

Funding for CDI projects varies from year to year and is directly influenced by the overall USGS budget. Since fiscal year 2009, CDI projects have been funded at a total of about $400,000 – $500,000 each year and it is anticipated that FY18 will be similar to previous years. Applicants can request funding up to $50,000 maximum and projects must include a minimum of 30% in matching funds.

Distribution of Funds

In FY18, CDI will distribute funds to the lead USGS Science Center only, and the lead Science Center will be in charge of further distributions to other Centers or external partners.

Estimated Schedule for Submission, Review, and Awards

RFP Information Session……January 16, 2018
Submission deadline for Statement of Interest……February 1, 2018
SOI Presentation Session……February 6, 2018
SOI Voting Closing……February 15, 2018

Applicants Notified and Full Proposals Requested……..……..……..February 26, 2018
Invited Full Proposals Due……..……..…....March 22, 2018
Funded Projects Announced……..……April 16, 2018
Awarded Funds must be spent……..……Sept 30, 2018
Final Project Reporting……March 1, 2019

Application Process

1. Submit Statement of Interest

Submit an SOI consisting of a 1-page narrative and a 1/2 page budget for initial evaluation. SOI guidance can be found in *Appendix A – Statement of Interest (SOI) Guidance.* This year the CDI will hold a session for submitters to present their proposals to the community in a lightning-talk format. The purpose of the session is to help submitters gain suggestions and support from the community, and help the community efficiently learn about the breadth of the SOIs. To aid the development of statements of interest, submitters may use the *FY18 RFP Collaboration Forum* to seek partners or other feedback.

2. Evaluate Statements of Interest

All SOIs will be reviewed by the CDI Community. The SOIs will be considered according to the *Evaluation Criteria for the Statement of Interest and Full Proposal*. Note that the CDI Executive Sponsors may select proposals that develop innovative methods for integrating, analyzing, or visualizing data in support of the USGS Acting Director's *Bureau priorities* or current topical emphasis areas, and a percentage of CDI Funds may be awarded to proposals that address those topics.

3. Invitation to Submit Full Proposal

Based on the SOI evaluation, selected SOI applicants will be invited to submit a Full Proposal. Full Proposal guidance can be found in *Appendix B – Invited Full Proposal Guidance.*

4. Full Proposal Review Process

Full Proposals will be evaluated according to the *Evaluation Criteria for the Statement of Interest and Full Proposal*. Proposals will be reviewed by a panel consisting of a professional peer group that is knowledgeable in data management, information technology, and other relevant disciplines in the context of the CDI. Recommendations by the Review Panel will be presented to the CDI Executive Sponsors for final selection.

Project Reporting

All CDI funded projects will be required to provide an informal mid-year briefing to the *CDI Facilitators* to communicate the status of the project. Projects must also contribute to the CDI Funded Project Report, which will be compiled in early 2019. At that time, project leads must provide a brief written report describing the project accomplishments, benefits, and deliverables with links to products or publications.

Description of the Request for Proposals

The CDI seeks to build and share knowledge about topics such as data integration, data handling and stewardship, scientific computing, and approaches for knowledge delivery. The main goal of CDI funding is to improve our collective knowledge about how to create better, longer-lasting and more accessible science products by leveraging the tools, methods and datasets available to the Earth and biological science communities. The CDI places high value on innovative projects that, in the near-term, produce new and reusable ideas, methods or tools that have an impact

beyond a single Program, Center, Region, or Mission Area. CDI project proposals will be evaluated based on the following guiding principles:

- Focus on targeted efforts that yield near-term benefits to Earth and biological science
- Leverage existing capabilities and data
- Implement and demonstrate innovative solutions (e.g. methodologies, tools, or integration concepts) that could be used or replicated by others at scales from project to enterprise
- Preserve, expose, and improve access to Earth and biological science data, models, and other outputs
- Develop, organize, and share knowledge and best practices in data integration

Unique Aspects of the CDI Request for Proposals Process

The CDI proposals process has several aspects that may be unfamiliar for first-time participants. The selection process incorporates community-involvement and multiple ways for submitters to communicate about their statements of interest.

All community members have the ability to comment and vote on the submitted Statements of Interest. Therefore, it is in the submitter's interest to promote their idea to the community, and to communicate the value added to the CDI by the proposed activities. This process may be puzzling to participants who are not accustomed to any promotion of proposals beyond submitting files to a review panel. However, the CDI views this as an opportunity to practice and improve plain-language communication and the ability to articulate the value of proposed activities. The community comment period also harnesses the expertise of the CDI community to make suggestions and improve on project ideas. These unique aspects help the CDI toward its goal of supporting the most useful and innovative ideas. In Phase 2 of the proposals process, after the community has given its input, a formal review panel evaluates the full proposals, similar to more typical proposal processes.

In recent years, the community has expressed that it is difficult to absorb the large number of ideas that are submitted in Phase 1 of the proposals process. This year, to assist in acquainting the CDI community with the submitted statements of interest, there will be an SOI Presentation Session at the beginning of the commenting and voting period. Each submitter will have the chance to present their SOI idea in an online one-minute lightning presentation. Community members will then have about a week to further investigate the statements and place their comments and votes.

FY18 Topical Emphasis - Risk assessment and hazard-exposure

Each year, the CDI has accepted statements of interest on any topic that follow the CDI guiding principles, noting that the CDI Executive Sponsors may select proposals that support the current Bureau Priorities. This year is no exception -- at the 2017 CDI Workshop, community members learned about the EarthMap vision, recently developed by the USGS Council of Senior Science Advisors (COSSA). EarthMap is an integrated scientific framework that spans traditional scientific boundaries and disciplines and integrates the full portfolio of USGS science (Jenni et al., 2017). A near-term focus of this vision was further refined in the *USGS FY18 Bureau Priorities* as a USGS Integrated Predictive Science Capacity that will "deliver powerful new products and services that provide: 1) vulnerability detection and assessment, 2) prediction and forecasting, 3) early warning, and 4) decision support at the scale of decisions." **This year, the CDI Executive Sponsors are encouraging proposals that produce building blocks for an Integrated Predictive Science Capacity in the specific focus area of risk assessment and hazard-exposure.**

We are still accepting submissions on any topic that follow the CDI guiding principles, but note that we are planning for a subset of supported projects this year to be related to risk assessment and hazard-exposure. Of total funding, CDI hopes to award approximately 50% of funding to qualified projects that are associated with the risk topic, depending on the number of relevant submissions received. Project teams in this focus area may be asked to work with the CDI facilitators to leverage related projects' work and progress.

Examples of risk-related projects include: integration of models and data to forecast invasive species, developing methods for aligning and integrating risk and hazard data from different sources, and technology to improve information delivery and stakeholder feedback in risk and hazard activities.

For more information about ongoing USGS Risk activities, please see the *Department of Interior Risk Assessment Project* document and the USGS Risk Research and Applications Plan. If the links to these resources are not available at the time of release of this guidance document, they will be posted on the *FY18 Proposals Wiki Page* at earliest availability.

In addition to the Bureau Priorities and the risk assessment focus, statements of interest that propose activities that address the recommendations from the 2017 CDI Workshop are encouraged. View the *Draft Workshop Proceedings* (link accessible to USGS and Dept. of Interior staff, email *cdi@usgs.gov* if you cannot access the document).

CDI Science Support Framework

Project proposals must also relate to elements of the *CDI Science Support Framework* (SSF), which categorizes and relates the activities and processes through which research data flows, and upon which the CDI operates. These elements include Data Management, Knowledge Management, the stages of the *Science Data Lifecycle Model*, Applications, Web services, Semantics, Information, Data assets, and Communities of Practice (*See*

*Appendix C – CDI Science Support Framework (*SSF)).

Examples of projects that relate to the goals and Science Support Framework

- Delivery of an immediate benefit to solve an existing data integration challenge, such as methods for blending datasets, or best practices for alignment/assimilation of data at different scales particularly with respect to *Bureau priorities*.
- Creation of innovative environments, tools, data stores, or services that enable discovery and usage of USGS data. This includes design patterns, management approaches, or products like web services or other software that can be used by other data publishers
- Development of standards or best practices for data management through community consensus building, such as convening a workshop and writing a white paper
- Development of a general ontology or tools for tagging data in support of standards and environments to facilitate discovery, understanding, and integration
- Testing or application of the aforementioned to a new, real-world problem to demonstrate and document strengths and issues for the purpose of feedback and improvement
- Exploitation of advanced or emerging technologies or approaches that enable new forms of USGS scientific knowledge creation or communication, such as developing mobile computing applications for rigorous data collection, or establishment of scientific policies or protocols around the novel component
- Development of innovative practices, methods, and strategies to better exploit collected data resources, such as data mining, parallel processing, large-scale data analysis, or scientific computing techniques and to improve data sharing, facilitate data preservation, and encourage lifecycle data management
- Development of vehicles to communicate or share knowledge, such as a committee to propose protocols/standards, workshops, online or in-person training course/materials, white paper, etc.
- Any of these or similar topics that have not been addressed otherwise within the agency

Proposal Concepts that should *not* be submitted to the CDI

- The CDI does not seek to supplant traditional natural science research or to fill a funding gap on a project supported elsewhere. Examples of topics that are a poor fit for CDI funding include:
- Supporting the collection of new data or field research.
- Monitoring, assessment, or dataset creation projects. Although the CDI may fund the creation of some broadly-usable ("foundational") data content, this is normally considered out of scope.
- Projects that would normally be funded by individual Program Areas.
- Projects that would normally be funded by other proposal processes such as the *John Wesley Powell Center for Analysis and Synthesis, Center of Excellence for Geographic Information Science (CEGIS)*, and *Office of Organizational and Employee Development (OED).*

Examples of past CDI Projects: *http://www2.usgs.gov/cdi/products-publications.html.*

Evaluation Criteria for the Statement of Interest and Full Proposal

Both the SOIs and Full Proposals will be evaluated based on the following six criteria. SOIs will only be expected to provide a concise statement in each of the criteria while Full Proposals must provide more detail. The evaluation weights (percentages) will only apply to the Full Proposal evaluation. For instructions on submitting SOIs, see *Appendix A – Statement of Interest (SOI) Guidance*; for Full Proposals, see *Appendix B – Invited Full Proposal Guidance*.

Scope (25%)

Evaluation will be based on whether the proposal adequately demonstrates the need for the effort/activity, how much the proposal contributes to the guiding principles and element(s) of the CDI Science Support Framework, and whether the effort has potential impact beyond a single Program, Center, Mission Area, or Region. CDI projects will also be evaluated on anticipated return on investment (e.g. cost savings, code utilization, publications, operational efficiencies, etc.).

Technical Approach (25%)

Evaluation will be based on the reasonableness of the technical approach applied to the problem and whether the approach is innovative or employs a proven, reliable technique that is appropriate to the problem. Evaluation will consider the steps, methodologies, technologies, and resources to be utilized in implementing the project. This includes facilities, computational/analytic platforms and tools, hardware/software, and other equipment supporting the project and/or its products.

Project Experience and Collaboration (25%)

Evaluation will be based on the appropriateness of the experience, special qualifications, and skills possessed by team members for successful completion of the proposed project. Evaluation will also consider whether inter-disciplinary or cross-Mission Area/Region collaboration and partnerships have been pursued where appropriate.

Sustainability, Outreach, and Communication (15%)

Evaluation will be based on how well the proposal describes the intended sustainability of the project deliverables (products, tools, services, metadata) for long-term access, reusability, and potential for integration, as well as the plan for communicating the value of the products during and after the project period. All products resulting from CDI projects must comply with the *Office of Science Quality and Integrity Instructional Memoranda* on data management. These products must be freely shared and made available, without charge or restriction, to the CDI, the broader USGS community, and beyond as appropriate. Software products developed with CDI funding must be uploaded to an appropriate code repository at the close of the funding period.

Budget Justification (5%)

Evaluation will be based on whether the budget is at or below $50,000 and meets the minimum 30% in-kind match. The budget should include travel to a relevant meeting. Because the CDI will not be hosting an in-person meeting in FY18, use an estimate for a meeting most relevant to your field. If you cannot decide on a meeting during the award period, use the *Earth Science Information Partners (ESIP) Summer Meeting* in San Diego, CA, from July 17-20, 2018. Evaluation will consider whether justification of salaries and contractor costs, travel, and equipment/publication costs are appropriate to project needs and the work hours proposed are reasonable within the timeframe. Projects with contractor support must describe how the contract work will be managed and documented to ensure that products are USGS property.

Timeline (5%)

Evaluation will be based on clear presentation of the project phases and milestones and the feasibility of the proposed workload given the short project duration. Although notification of award may come earlier, assume that funding will be awarded no sooner than May 31, 2018 and reference specific months or dates within FY18 or relative to time from date of award (e.g., 3 weeks after award date). The timeline must demonstrate reasonable completion and complete use of funds by September 30, 2018. Recognizing the USGS publication process may take additional time beyond the end of the project, indicate the anticipated publication date for any USGS publications resulting from the project.

Appendix A – Statement of Interest (SOI) Guidance

1) Prepare a Statement of Interest document

Prepare a Statement of Interest document that includes the following information:

- Project title and name of the USGS lead PI
- Project narrative, addressing the topics in the *Evaluation Criteria for the Statement of Interest and Full Proposal (***1-page maximum)**
- Estimated budget table **(use template provided below, 0.5 page)**

Use Times New Roman 11 point font with one-inch margins, and save the file (which should be 1.5 pages total) as a PDF document.

Estimated Budget Table (½ page)

Budget Category	Federal Funding "Requested"	Matching Funds "Proposed"
1. PERSONNEL (SALARIES including benefits):		
Federal Personnel Total:	$	$
Contract/Collaborator Personnel Total:	$	$
Total Salaries:	$0	$0
2. TRAVEL EXPENSES:		
Travel Total (Per Diem, Airfare, Mileage/Shuttle) **x** # of Trips:	$	$
Other Expenses (e.g. Registration Fees):	$	$
Total Travel Expenses:	$0	$0
3. OTHER DIRECT COSTS: (itemize)		
Equipment (including software, hardware, purchases/rentals):	$	$
Publication Costs:	$	$
Office Supplies, Training, Other Expenses (specify):	$	$
Total Other Direct Costs:	$0	$0
Total Direct Costs:	$0	$0
Indirect Costs (%):	$0	$0
GRAND TOTAL:	$0	$0

**Note: Travel must be included for at least one representative to attend a relevant meeting. Travel cannot include data field collection. See the budget justification section of the Evaluation Criteria for the Statement of Interest and Full Proposal for more information.*

2) Register on the Online Proposal Management System

Register on the Online Proposal Management System (*https://my.usgs.gov/CDI_RFP*) and upload your SOI file. The system will be open by early January 2018. Log in with your USGS Active Directory username and password. The proposal management system collects all administrative information, that is why it is not necessary to include anything other than the project title and name of the USGS lead PI on the actual SOI document. Please respond to all questions in the proposal management system or the submission will be considered incomplete and invalid.

Required Fields in the Online Proposal Management System

We suggest that you prepare this information before registering on the online proposal management system.

- **Title:** Include a descriptive title of the proposed project.
- **Principal Investigator(s):** List the lead USGS Principal Investigator with affiliation (Mission Area, Region,

Organization), ORCiD (if applicable), email, and location, as well as the same contact information for any co-Principal Investigators. **NOTE**: If you are submitting an SOI on behalf of the lead PI, you will be identified as the "Registrant" from your log-in credentials. Make sure your email is listed in the "Registrant E-Mail" and the PI's email is in the "PI E-Mail" field. As a default, future notifications about the submission will go to the registrant email unless you get in contact with us for a different arrangement.

- **Collaborators:** Provide the names, affiliation (Mission Area, Region, Organization), ORCiD (if applicable), email, and location for other project personnel involved.
- **CDI Science Support Framework Element(s):** Indicate up to three element(s) of the CDI SSF that the proposal covers (See *Appendix C - CDI Science Support Framework (SSF)*).
- **Lead Cost Center:** The single Cost Center where funds will be distributed (usually the Cost Center of the lead USGS Principal Investigator)
- **Project Description:** 1-2 sentences used to briefly describe your SOI beyond the title. **(2 sentence maximum)**
- **List of anticipated deliverables from the project:** Include a brief list of the types of products that will be generated as a result of the project, e.g. mobile application, fact sheet, GIS shape file, GIS data layer, desktop data entry application, online data entry application, online data cleaning application, USGS Blog article or press release, etc. The list should consist of short phrases as demonstrated in the previous sentence.
- **Total budget:** Provide the total project funding requested from CDI, matching the value on your estimated budget table, not to exceed $50,000.
- **In-Kind Match:** Provide the total in-kind funding from other sources, matching the value on the estimated budget. There must be a minimum of 30% in-kind match.

Appendix B – Invited Full Proposal Guidance

Proposals must be submitted through the online proposal management system at *https://my.usgs.gov/CDI_RFP*. The proposal management system has collected all administrative information in Phase 1 with the SOI, so it is not necessary to include a cover sheet page with the full proposal. However, please be sure to complete any questions that appear within the online proposal management system and update any responses that may have changed from Phase I for your proposal to be considered complete and valid.

1) Prepare three separate documents to the online proposal management system.

The Budget Form and Data Management Planning form are available on the *CDI FY18 RFP Forms* page.

- ✓ A **Full Proposal, single PDF document** (not to exceed 10MB) with:
  - General Public Summary (not to exceed 150 words; submitted on a separate page within the proposal)
  - Proposal Narrative (max. 7 pages)
  - Appendices (e.g., CVs – max. 2 pages each, letters of support – max. 1 page each)
- ✓ A **Budget Form** using the MS Excel template
- ✓ A **Data Management Planning Form, single PDF document (**using the MS Word template**)**

Proposals should be formatted to standard letter size (8.5" W by 11" L). All proposals should be no more than 7 pages, single-spaced, not including the general public summary and appendices. Narrative (body) text must be rendered in Times New Roman 11 point font, excluding headings which must be formatted bold and 12 point. All pages (including appendices) must be numbered. Failure to follow the stated guidelines may reflect negatively on the proposal.

All graphics, photos, illustrations, tables, graphs, and charts must be embedded directly in the proposal document and be specifically referenced at least once in the narrative (body) of the proposal. All graphics must be accompanied by a caption that describes the graphic. These count towards the total number of pages allotted.

Full Proposal Template

A. General Public Summary (on a separate page)

Provide a synopsis of the overall project that is written for a general public audience, does not exceed 150 words, and is suitable for sharing on public Web sites and other outreach methods. Key points to include: Why is the project important? Why should the public care? What will the project accomplish? How will the results of the project work toward the goals of CDI and help stakeholders?

The general public summary should be submitted on a separate page within the proposal PDF document.

B. Proposal Narrative (max. 7 pages)

The main body of the proposal should consist of six sections which will be evaluated by the Review Panel:

- Scope
- Technical Approach
- Project Experience and Collaboration
- Sustainability, Outreach, and Communication
- Budget Justification
- Timeline

Prepare the sections according to the *Evaluation Criteria for the Statement of Interest and Full Proposal.*

Additional information specific to the Full Proposal phase are noted below.

- Budget Justification: Proposals may not exceed $50,000 in requested funding and must include a minimum of 30% in-kind match within the overall budget. **The lead PI(s) must work with their Administrative Officer (AO) to ensure an accurate budget and funding management responsibilities before submission.** Proposals utilizing USGS contracting staff must include in the Budget Justification statement a confirmation from the Contracting Officer's Representative (COR) that there is an available contract to complete the project. All CDI funds will be transferred to the lead USGS cost center through a change of allocation. The lead USGS cost center may then provide sub-awards to other collaborating organizations/cost centers.

  Full Proposals must submit a *Budget Form* and include the Budget Justification statement to explain project costs in the following categories:
  - *Personnel (Salaries including benefits):* Include estimates (by hours) and rate of compensation proposed for each named individual or category (e.g., graduate student). Ensure that the identified personnel and their affiliations are clearly listed. Projects with contractor support must describe how the contract work will be managed and documented to ensure that products are USGS property.
  - *Travel Expenses:* Specify travel requirements for project meetings, and/or conference attendance. Itemize estimated travel costs to show the number of trips required, destinations, the number of travelers and per diem rates, cost of transportation (e.g., vehicle rental), and miscellaneous expenses for each trip. Travel must be included for at least one representative to attend a relevant meeting. Travel cannot include data field collection. Because the CDI will not be hosting an in-person meeting in FY18, use an estimate for a meeting most relevant to your field. If you cannot decide on a meeting during the award period, use the *Earth Science Information Partners (ESIP) Summer Meeting* in San Diego, CA, from July 17-20, 2018.
  - *Other Direct Costs:* Itemize any proposed permanent equipment acquisitions ($5,000 or more) and show each estimated cost. Explain costs including publication costs, office supplies, training, etc.
  - *Indirect Costs (Overhead):* Provide indirect cost rate and amount approved for each institution.

C. Appendices

- Required: CV(s) of Principal Investigator(s) that highlights relevance to the proposed work **(max. 2 pages each)**
- Optional: CV(s) of other collaborator(s) that highlights relevance to the proposed work **(max. 2 pages each)**
- Optional: Letters of support from USGS or outside partners indicating a clear need for this effort. Submissions may also include Memoranda of Understanding (MOU) and/or letters of support indicating commitment to the longevity of the project. **(max. 1 page each)**

Budget Form

Applicants are required to use the *Budget Template* (in MS Excel format). Include Personnel, Travel Expenses, and Other Direct Costs, separating the CDI funds from the in-kind match as indicated in the template at the link above.

Data Management Planning Form

Applicants are required to use the *Data Management Planning (DMP) Form Template* and submit in PDF format. This year, the DMP template is new and streamlined. Please do not use a template from a previous year. The

information requested on the Data Management Planning Form helps project teams plan for data management and product communication needs. For more guidance on data management plans, see the *USGS Data Management Website*, specifically the *Data Management checklist*. All products resulting from CDI projects must comply with the *Office of Science Quality and Integrity Instructional Memoranda* on data management.

2) Update fields in the Online Proposal Management System and submit your three documents

Log on to the Online Proposal Management system (*https://my.usgs.gov/CDI_RFP*) and make sure your information is correct. Upload the full proposal, budget, and data management plan into the system.

A cover sheet will be generated by the proposal management system. The cover sheet will aid reviewers and the review process by allowing them to easily distinguish between proposals and see each proposal's basic elements at a glance. The proposal management system has collected all administrative information in Phase 1 of the RFP, so it will not be necessary to re-enter the information or include a cover sheet page with the full proposal. **However, make sure to update any fields that have changed since Phase 1, especially the budget totals and anticipated deliverables.**

Appendix C – CDI Science Support Framework (SSF)

The Community for Data Integration (CDI) represents a dynamic aggregation of multiple communities of practice, focused on the advancement of scientific data and information management and integration capabilities across the USGS and external organizations.

Since 2009, CDI has funded a variety of projects that support the overarching goal of data integration. USGS and other researchers conduct monitoring, assessment, and research activities that generate data assets. Through the application of business, computational, and analytic processes and technologies, these data assets are converted into information that contributes to our understanding of the Earth's physical and biological systems. This is the context within which data management and integration occur and where the CDI operates (Fig. 1).

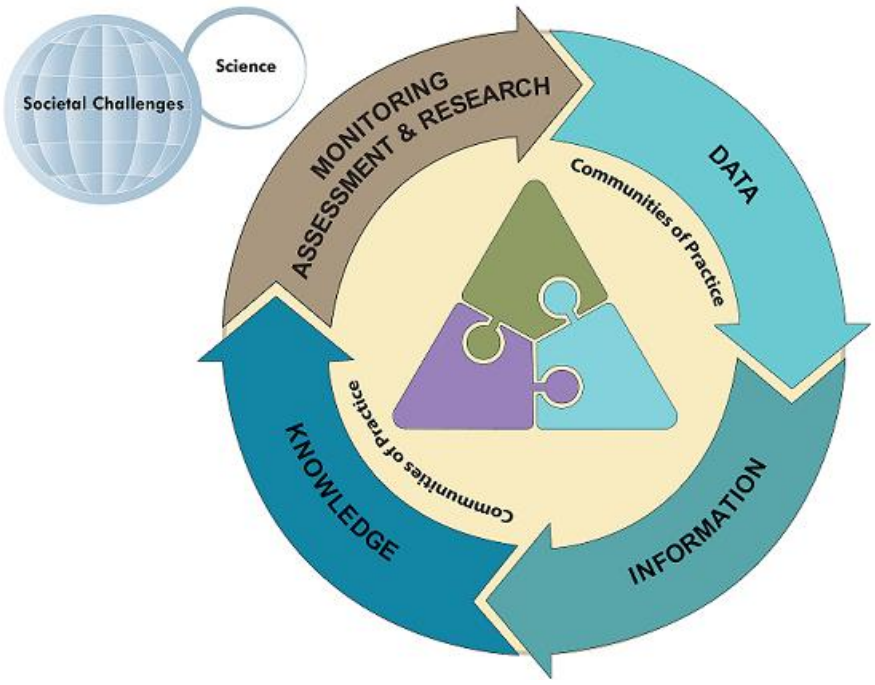

Figure 1: Overview of CDI Operational Context

Communities of Practice	Communities of practice include scientists, the CDI as a whole, CDI Working Groups, external partners, and the human network of scientific domain collaborators.
Computational Tools & Services	Computational tools and services include applications, Web services, data discovery tools, models, semantic services and tools, infrastructure, data brokers, and visualization tools.
Management, Policy & Standards	Management, policy, and standards include data stewardship, the implementation of the Science Data Lifecycle, knowledge management, data standards, governance, and policy.
Data & Information Assets	Data and information assets include persistent archives, data registries, catalogs, data, metadata, derived information products, knowledge bases, and vocabularies/ontologies.

The CDI SSF (Fig. 2) provides a conceptual architecture that illustrates how the CDI contributes to Bureau-level data integration efforts; and defines how current and future CDI projects fit within the *framework*.

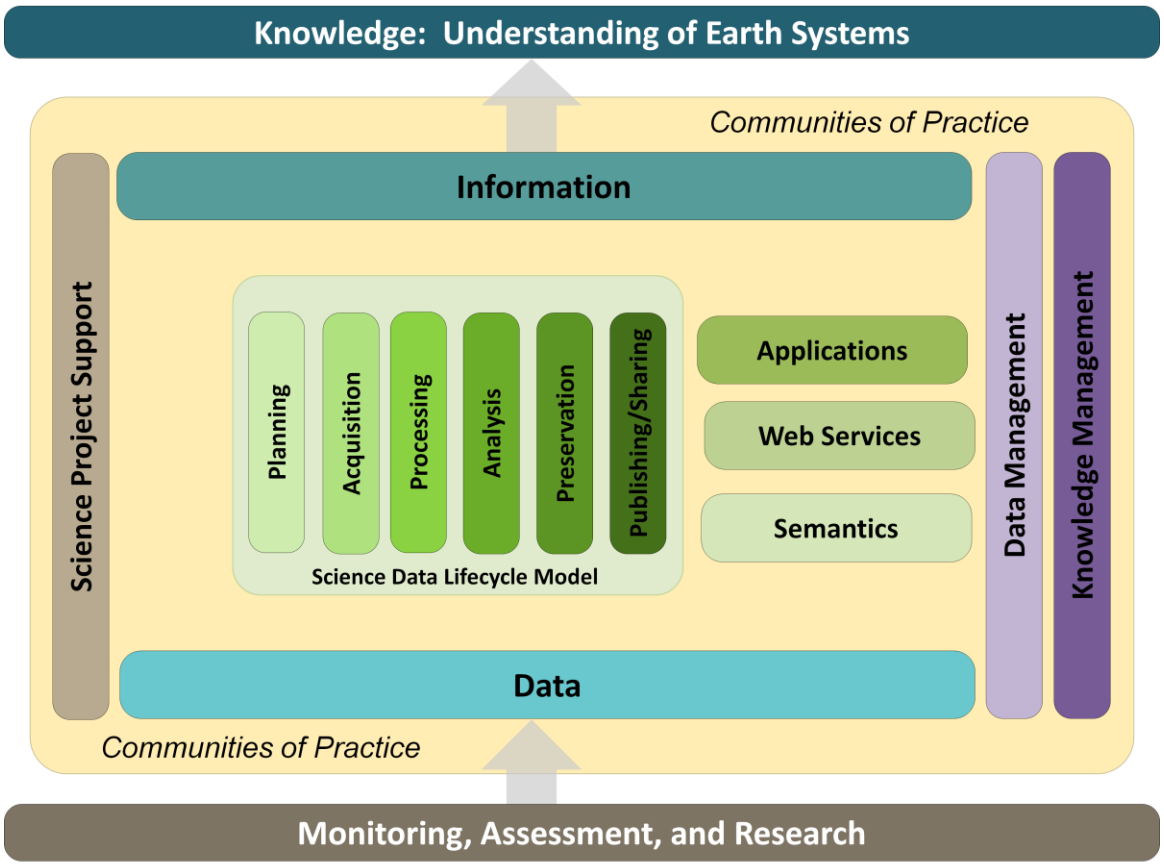

Figure 2: CDI Science Support Framework

USGS Data Assets Flow through the CDI Science Support Framework. USGS data assets flow vertically through the SSF from a base of monitoring, assessment, and research through the Science Data Lifecycle, applications, Web services, and semantics. The assets are transformed into information products that benefit from data and knowledge management and also increase knowledge and understanding of the Earth's physical and biological systems. Data assets flow horizontally through the SSF from science project support to data and knowledge management.

The horizontal elements in the SSF represent the "what" of the CDI: products and tools that contribute to the advancement of scientific data and lead to the development of knowledge and understanding of the Earth's systems.

The vertical elements in the SSF represent the "how" of the CDI: the processes, the implementation of standards and best practices, and the interactions among people, data, and technology used to achieve data integration.

Individual Framework element descriptions:

Science Inputs (brown elements)

***Monitoring, Assessment, & Research*:** USGS scientists conduct monitoring, assessment, and research that generate data assets. Through the application of business, computational, and analytical processes and technologies, these assets are converted into information

products that can be shared with other researchers, stakeholders, and citizens to increase our knowledge and understanding of the Earth's physical and biological systems.

***Science Project Support*:** Successful science projects encompass a range of activities represented in the Data Lifecycle. At each step in the cycle, researchers and data stewards rely on an array of sophisticated tools and services for data, information and knowledge discovery, acquisition, integration, management, and sharing.

Communities of Practice (tan element)

Communities of practice are the foundation for CDI and all its products – the communities of people working towards the goal of advancing scientific data and information management and data integration across the USGS.

Data & Information Assets (blue elements)

USGS assets include **Data** (e.g., raw data, databases, and linked open data (RDF[1])); **Information** or derived/interpreted information products (e.g., published or shared maps, reports, datasets); and **Knowledge** of all types and in all forms — recorded, organized, and preserved in the form of artifacts. Knowledge can be improved, shared across groups, organizations, and domains, and reused to support learning and research.

Computational Tools & Services (green elements)

Science Data Lifecycle include tools and services that move data through the lifecycle, human and machine interactions, and interactions with data through technology.

Detailed descriptions of the Science Data Lifecycle:

- **Planning** – A documented sequence of intended actions to identify and secure resources and gather, maintain, secure, and utilize data assets.
- **Acquisition** – The series of actions for collecting or adding to data assets.
- **Processing** – A series of actions or steps performed on data to verify, organize, transform, integrate, and extract data in an appropriate output form for subsequent use.
- **Analysis** – A series of actions and methods performed on data that help describe facts, detect patterns, develop explanations, and test hypotheses.
- **Preservation** – Actions and procedures to keep data for some period of time; to set data aside for future use.
- **Publishing/Sharing** – To prepare and issue, or to disseminate data or information products.

Semantics convert raw data into data that can be interpreted by machines: Machine Readable Metadata, Semantic Mediation for Data Integration & Discovery, Ontologies/Vocabularies, and World Wide Web Consortium Standards.

Web Services include machine to machine data exchange, SOAP,[2] REST,[3] SPARQL[4] EndPoints, and other protocols and services.

Applications include human readable data services and user interfaces to data driven applications.

Management, Policy, & Standards (purple elements)

Data Management includes data and metadata standards and policies and occurs in all phases of the Data Lifecycle from scientific research to finished information products.

Knowledge Management involves the creation, standardized documentation, and organization of knowledge using tools such as SKOS[5] Vocabularies and information modeling, resulting in the formation of knowledge bases.

[1] Resource Description Framework
[2] Simple Object Access Protocol
[3] REpresentational State Transfer
[4] SPARQL Protocol and RDF Query Language
[5] Simple Knowledge Organization System

Appendix D – CDI Coordinators

We encourage proposers to get in touch with relevant CDI contacts to discuss their proposals.

CDI Executive Sponsors
Kevin Gallagher, Associate Director, USGS Core Science Systems
Tim Quinn, Associate Chief Information Officer, Office of Enterprise Information
Cheryl Morris, Director, USGS Core Science Analytics, Synthesis and Libraries Program

CDI Facilitators
Leslie Hsu (lhsu@usgs.gov)
Madison Langseth (mlangseth@usgs.gov)

Collaboration Area leads may have specific suggestions or contacts for their focus topic. All CDI Collaboration Area pages can be accessed at *https://my.usgs.gov/confluence/x/yhv1I*.

Selected CDI Collaboration Area Leads:

Bioinformatics
Denise Akob, Robert (Scott) Cornman, and Christina Kellogg

Data Management
Vivian Hutchison and Cassandra Ladino

Data Science
Lindsay Carr

Earth-Science Themes
Roland Viger

Metadata Reviewers
Fran Lightsom

Open Source
Cassandra Ladino

Semantic Technologies
Fran Lightsom

Software Development
Michelle Guy and Blake Draper

Subduction Zone
Joan Gomberg

Tech Stack
Richard Signell

Appendix E - Additional Instructions for Project Products

Use the specific wording below to acknowledge funding in CDI publications and products:

This work was supported by funding from the USGS Community for Data Integration (CDI).